

**ADDRESSING TRACK COALESCENCE IN  
SEQUENTIAL *K*-BEST MULTIPLE HYPOTHESIS  
TRACKING**

A Thesis  
Presented to  
The Academic Faculty

By  
Ryan D. Palkki

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science  
in  
Electrical and Computer Engineering



School of Electrical and Computer Engineering  
Georgia Institute of Technology  
August 2006

# **ADDRESSING TRACK COALESCENCE IN SEQUENTIAL $K$ -BEST MULTIPLE HYPOTHESIS TRACKING**

Approved by:

Dr. Aaron Lanterman, Advisor  
*School of Electrical and Computer Engineering  
Georgia Institute of Technology*

Dr. James McClellan  
*School of Electrical and Computer Engineering  
Georgia Institute of Technology*

Dr. Magnus Egerstedt  
*School of Electrical and Computer Engineering  
Georgia Institute of Technology*

Date Approved: May 10, 2006

*To my beautiful wife, Brenna.*

## **ACKNOWLEDGMENTS**

I wish to thank my advisor, Aaron Lanterman, for his ongoing guidance over this work. I also wish to thank Dale Blair for his support and suggestions throughout the course of this research. Finally, I wish to thank the other members of my group for all the time and help that they are always willing to give.

This work was funded by U.S. Office of Naval Research grant N00014-04-C-0406.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b>	iv
<b>LIST OF TABLES</b>	vii
<b>LIST OF FIGURES</b>	viii
<b>SUMMARY</b>	ix
<b>CHAPTER 1 INTRODUCTION</b>	1
1.1 The Data Association Problem	1
1.2 Nearest Neighbor and Track Splitting Filters	1
1.3 The Optimal Filter and the Criteria for the Estimator	2
1.3.1 “Hard” (MAP) vs. “Soft” (MMSE) Association	2
1.4 Approximations to the Optimal Filter	3
1.4.1 The PDA and $N$ -Scan Approximations to the Optimal MMSE Filter	4
1.4.2 Sequential $K$ -Best MHT	5
1.5 Outline of the Rest of the Thesis	8
<b>CHAPTER 2 SIMULATION FRAMEWORK</b>	9
2.1 Assumptions and Filter Model	9
2.2 Track Score Function	10
2.3 Reasonability of the MAP estimate for $K$ -Best MHT	11
2.4 Performance Metrics	12
2.5 Truth Trajectory, False Alarms, and Gating	13
2.6 Process Noise Tuning	14
<b>CHAPTER 3 COMPARING <math>K</math>-BEST MHT WITH PDA</b>	18
<b>CHAPTER 4 DEALING WITH TRACK COALESCENCE</b>	20
4.1 The Simple “Measurement Constraint” Method	20
4.2 Tighter Coalescence Tests	21
4.2.1 Tighter Test I	21
4.2.2 Tighter Test II	22
4.3 Looser Coalescence Tests	24
4.4 The J-Divergence Test	26
<b>CHAPTER 5 CONCLUSION</b>	30
<b>APPENDIX A RESULTS OF “TIGHTER TEST I” OF SECTION 4.2.1</b>	31
<b>APPENDIX B RESULTS OF THE MAHALANOBIS TESTS OF SECTION 4.3</b>	32

<b>APPENDIX C KULLBACK-LEIBLER DISTANCE FOR MULTIVARIATE GAUSSIAN NORMAL DISTRIBUTIONS . . . . .</b>	<b>33</b>
<b>REFERENCES . . . . .</b>	<b>36</b>

## LIST OF TABLES

Table 1	Process noise parameters tested. . . . .	15
Table 2	Which $\sigma_v^2$ works best? . . . . .	16
Table 3	Track loss ratio of PDA. . . . .	18
Table 4	Track loss ratio of $K$ -Best MHT (with measurement constraint to prevent track coalescence). . . . .	18
Table 5	The bottom left corner of Table 4, but now with 10,000 Monte Carlo Runs. . . . .	19
Table 6	Track loss ratio of $K$ -Best MHT without any method to prevent track coalescence; compare with Table 4. . . . .	21
Table 7	Track loss ratio of $K$ -Best MHT with the J-Divergence Test. . . . .	29
Table 8	Track loss ratio of $K$ -Best MHT when the threshold is 2. . . . .	31
Table 9	Track loss ratio of $K$ -Best MHT when the threshold is 3. . . . .	31
Table 10	Track loss ratio of $K$ -Best MHT when the threshold is 4. . . . .	31
Table 11	Track loss ratio of $K$ -Best MHT when $d_1^2$ is used as the Mahalanobis distance. . . . .	32
Table 12	Track loss ratio of $K$ -Best MHT when $d_2^2$ is used as the Mahalanobis distance. . . . .	32
Table 13	Track loss ratio of $K$ -Best MHT when $d_3^2$ is used as the Mahalanobis distance. . . . .	32

## LIST OF FIGURES

Figure 1	Flowcharts of the NN and $K$ -Best Trackers. . . . .	6
Figure 2	Results after each of the six steps of the $K$ -Best algorithm as numbered in the flowchart of Figure 1(b). In this example, $K = 2$ , and we happen to also have two measurements. For simplicity, gating is not shown, and the missed-detection hypotheses are not considered (both of these steps are straightforward). $L$ is the cumulative track score, $\Delta L$ is the incremental score, ‘+’ denotes a track prediction, ‘*’ represents a measurement, and the connecting lines are the possible data associations. The stars are filtered estimates, and the filled star is the output track state estimate of the $K$ -Best tracker. . . . .	7
Figure 3	Illustration of the Sequential $K$ -Best MHT tracker. Target is moving from right to left. . . . .	8
Figure 4	Truth trajectory and measurements of a simulation. Target is moving from right to left. . . . .	14
Figure 5	Histogram of the results presented in Table 2. The height of the graph at any of the seven process noises is the number of tracking scenarios for which that process noise was preferred over all of the others. . . . .	17
Figure 6	Track loss ratio for the $P_D = 0.94, \beta_{FA} = 0.04$ case. . . . .	19
Figure 7	Illustration of track coalescence when $K = 3$ . Target is moving from right to left. . . . .	20
Figure 8	Track loss ratio when $P_D = .94, \beta_{FA} = .04$ . . . . .	23
Figure 9	Track loss ratio when $P_D = .94, \beta_{FA} = .04$ . . . . .	24
Figure 10	Track loss ratio when $P_D = .94, \beta_{FA} = .04$ . . . . .	26
Figure 11	Even if two tracks are close together, they are likely to pick different measurements if the covariances are different. . . . .	27
Figure 12	Track loss ratio when $P_D = .94, \beta_{FA} = .04$ . . . . .	29



## SUMMARY

Multiple Hypothesis Tracking (MHT) is generally the preferred data association technique for tracking targets in clutter and with missed detections due to its increased accuracy over conventional single-scan techniques such as Nearest Neighbor (NN) and Probabilistic Data Association (PDA). However, this improved accuracy comes at the price of greater complexity. Sequential  $K$ -best MHT is a simple implementation of MHT that attempts to achieve the accuracy of multiple hypothesis tracking with some of the simplicity of single-frame methods.

Our first major objective is to determine under what general conditions Sequential  $K$ -best data association is preferable to Probabilistic Data Association. Both methods are implemented for a single-target, single-sensor scenario in two spatial dimensions. Using the track loss ratio as our primary performance metric, we compare the two methods under varying false alarm densities and missed-detection probabilities.

Upon implementing a single-target Sequential  $K$ -best MHT tracker, a fundamental problem was observed in which the tracks coalesce. The second major thrust of this research is to compare different approaches to resolve this issue. Several methods to detect track coalescence, mostly based on the Mahalanobis and Kullback-Leibler distances, are presented and compared.

# CHAPTER 1

## INTRODUCTION

### 1.1 The Data Association Problem

Assuming that a good linear dynamics model for a target of interest is available, along with a sequence of measurements that are linearly related to the target state and that are corrupted by Gaussian noise, the target tracking problem is well solved by the Kalman Filter. The Kalman Filter is the optimal recursive data processing filter in such cases.<sup>1</sup> However, the classic Kalman Filter equations do not take into account any *uncertainty* in the origins of the measurements; they assume *we know* that the measurement received on each scan originated from the target of interest [1]. The Kalman Filter deals optimally with measurement error, but it does not deal at all with measurement-origin uncertainty. In a realistic tracking scenario, there will be measurements originating from other targets or from clutter (false alarms), and sometimes no measurement will be received at all even though the target of interest was there (missed detections). The Kalman update equation may then use the wrong measurement to update the track estimate, and the tracker performance may be severely degraded. For this reason, the problem of “data association,” i.e., the problem of deciding which—if any—measurement originated from which target, has historically been one of the greatest problems in target tracking.

### 1.2 Nearest Neighbor and Track Splitting Filters

The simplest method of data association is the Nearest Neighbor filter. This method finds the measurement from the current scan that is statistically closest to the track prediction. It then uses this “nearest neighbor” in the Kalman Filter update equation, while ignoring all of the other measurements from that scan. Since this method may ignore some of the observed data, it is reasonable that better data association algorithms should be possible.

---

<sup>1</sup>If the dynamics model and/or the measurement model are nonlinear, an Extended Kalman Filter (EKF) may be employed. The EKF algorithm is suboptimal, but it is effective in many applications.

One of the earliest data association algorithms was the track splitting filter [1], also called the branching algorithm [2], which simply splits the track whenever multiple measurements fall within the validation (gating) region<sup>2</sup> of the track prediction. The number of branches of the track can grow exponentially with time, and it becomes necessary to “prune” (eliminate) the tracks whose likelihood is below a threshold. The number of tracks (and the memory requirements) still grow with time, despite this pruning.

### 1.3 The Optimal Filter and the Criteria for the Estimator

Understanding the track splitting filter helps in thinking about the general problem of data association. If the track splitting filter were to run for  $k$  scans, and there was no pruning of unlikely tracks, then there would be a large number of possible tracks by the time that scan  $k$  is received (if  $M$  is the number of possible tracks, then  $M$  grows exponentially with  $k$ ). Each of those  $M$  tracks has a certain probability of being the true track. For any individual track  $j$  among those  $M$  possible tracks, *given* that track  $j$  is the true track, the posterior pdf of the target state would be Gaussian. Since it is not known which of the  $M$  branches is the true target track, the posterior pdf of the target state given *all* of the measurements from all of the scans will be a probabilistically weighted mixture of Gaussians.

It is theoretically possible to keep track of all of the possible tracks (i.e., to not prune any of the branches, no matter how unlikely) and to compute the probability of each one. It is thus also possible, at least in theory, to compute the Gaussian mixture posterior distribution of the target state given all received measurements over the entire history of scans through scan  $k$ .

#### 1.3.1 “Hard” (MAP) vs. “Soft” (MMSE) Association

Once the posterior distribution of the target state is found, one can take its mean to find the minimum-mean-squared-error (MMSE) estimate of the target state. This is a “soft

---

<sup>2</sup>To reduce computational requirements, any measurements that fall outside of a “gate” region centered around the track prediction are not considered in the data association problem.

association” approach; the estimate used for the track update does not correspond to one particular measurement, but rather results from a weighted combination of data associations from all of the possible measurements. Although this estimate is optimal in the MMSE sense, there are situations in which it is desirable to make a hard decision rather than a soft decision. As Drummond [3] points out, “for estimation involving hypotheses due to discrete possibilities, the traditional MMSE criterion leads to so called soft decisions that may not be appropriate for an interceptor with a small region of lethality while, in contrast, hard decisions might increase the probability of kill.” For this reason, the maximum a posteriori (MAP) estimate, which finds the state that maximizes the posterior distribution, may be preferred. Most Multiple Hypothesis Tracking (MHT) methods are essentially MAP estimators [4]. In section 2.3, we show that a hard (MAP) association approach appears to be reasonable for our Sequential  $K$ -Best MHT.

## 1.4 Approximations to the Optimal Filter

The optimal<sup>3</sup> Bayesian (MMSE) approach described above was implemented for a simple scenario by Singer et al. [5] in 1974. The major problem with this method (as well as *any* method that employs no scheme to prune and manage the number of tracks, including the optimal MAP method) is that it is practically infeasible for realistic scenarios. As mentioned previously, the number of branches grows exponentially with time. So, instead of implementing the full-blown optimal algorithm, most classical tracking methods aim for one of the optimal criteria (MMSE or MAP, or something else, like maximum likelihood (ML)), but impose some approximations to make the algorithm feasible. For example, the track splitting filter becomes feasible when the unlikely branches are eliminated. Here, the approximation made to the optimal filter is done by representing all of the possible

---

<sup>3</sup>The approach is “optimal” among methods that assume that multiple measurements will not originate from a single target; so far, we have made this common assumption. If there is one track and  $N$  measurements under consideration, this constraint cuts the  $\sum_{n=0}^N \binom{N}{n} = 2^N$  association hypotheses down to  $N + 1$  hypotheses per scan. After  $M$  scans, there will be  $(N + 1)^M$  tracks, compared to  $(2^N)^M$ . While the number of branches still grows exponentially with the number of scans, it grows at a much slower rate. In comparison, the single-target  $K$ -Best MHT maintains a constant number of tracks and the load remains fixed.

branches by a subset of the most likely branches. With this approximation, it is feasible to compute an approximate posterior distribution. If the goal is to minimize MMSE, then a soft association will be used to obtain the track state estimate from the (approximate) posterior distribution. If a MAP or ML criterion is used, then a hard association will be made according to that criterion.

#### 1.4.1 The PDA and $N$ -Scan Approximations to the Optimal MMSE Filter

The Probabilistic Data Association (PDA) Filter, sometimes referred to as a “suboptimal Bayesian approach,” is another example of a suboptimal algorithm. PDA<sup>4</sup> seeks to minimize MMSE; hence, it is a soft association approach. It makes a major approximation to the optimal MMSE filter: whereas the posterior density of the state given all past observations is really a Gaussian mixture, the PDAF approximates this density with a single Gaussian with the same mean and variance [6]. Using PDA, only one track is propagated forward at each scan (instead of an exponentially increasing number of tracks), and the computational requirements do not increase with time. Because it takes all of the measurements from the current scan into account, it performs better in dense clutter than the NN filter. However, the extra computational burden is small (“about 50% greater” than the NN Filter [6]).

When Singer et al. derived and implemented the optimal Bayesian (MMSE) filter [5], they also proposed an approximation to it. Instead of splitting the track all the way back through its history, they split it for only the last  $N$  scans. This approximation is equivalent to combining any tracks that have the last  $N$  measurements in common. Since this suboptimal version still seeks the MMSE criterion, it uses the soft association approach and recombines the tracks into one mixed estimate. Thus, when  $N = 0$ , this so-called Generalized Pseudo-Bayesian filter [4] reduces to the PDA filter. Singer et al. then show that  $N = 1$  and  $N = 2$  gave “near-optimal” results [7]. Note that this  $N$ -scan method requires a history, i.e. “window,” of data if  $N > 0$ . Improved performance comes with the cost of greater

---

<sup>4</sup>The Joint Probabilistic Data Association (JPDA) method is an extension of PDA to the multiple-target case.

complexity.

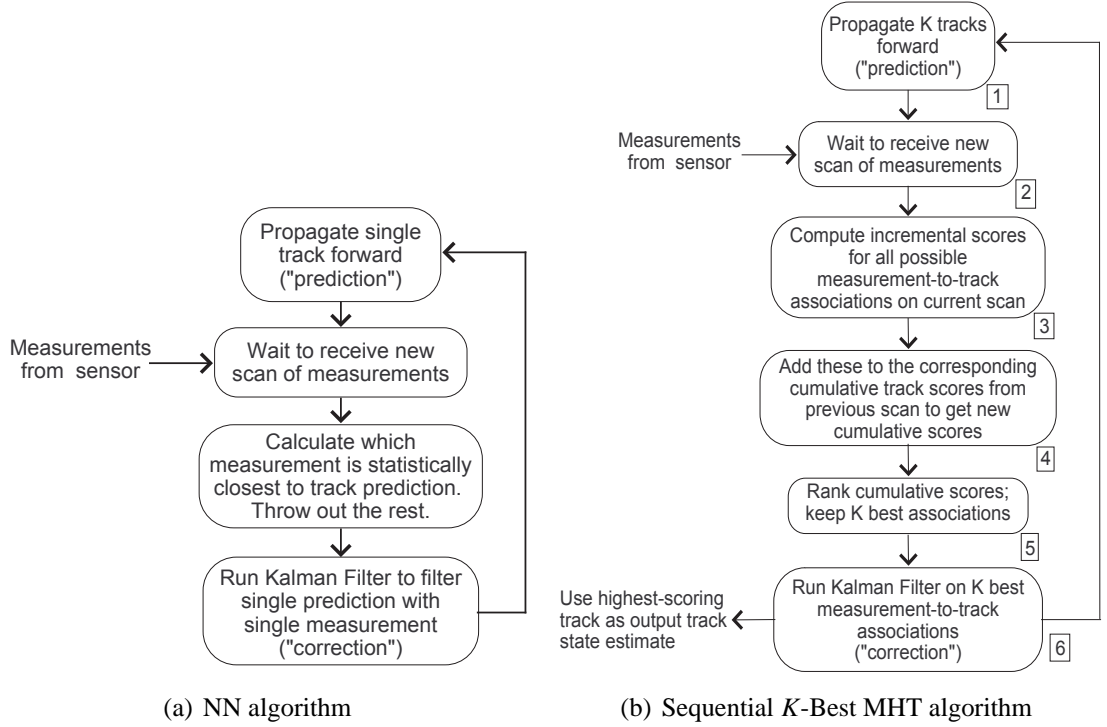
#### 1.4.2 Sequential $K$ -Best MHT

The above examples illustrate the nature of most classical tracking methods. The theoretically optimal algorithms described previously are infeasible, so approximations are made to make them feasible. This is also the case with the Sequential  $K$ -Best Multiple Hypothesis Tracker.

The Sequential  $K$ -Best MHT is similar to the track-splitting filter, except that it allows only  $K$  total tracks to exist at each scan. The  $K$ -Best method finds the  $K$  best measurement-to-track assignments (given by the recursive probability-based score of [8]; see Section 2.2) for each scan of data received. The  $K$  hypotheses are ranked from the most likely down to the  $K^{th}$  most likely hypothesis. Then, on the next scan, it finds the  $K$  best associations from all of the  $K$  tracks maintained from the previous scan. At each scan, the track with the highest score is the output track state estimate (see Section 2.3). This process is illustrated and compared with the Nearest Neighbor method in Figure 1. The six steps of the algorithm are illustrated for a simple example in Figure 2.

Figure 3 illustrates how the  $K$ -best MHT tracker “feels around for” the true trajectory. As shown in the figure, the best track at the current scan is sometimes later shown not to be the best when considered in the light of further data. In this way, Sequential  $K$ -Best MHT avoids losing the track as NN would have.

Another difference between NN and  $K$ -Best MHT is that while the NN filter always updates the track with the nearest measurement, the MHT filter also considers the case that all the measurements may be false alarms, i.e., there is a missed detection. Sometimes, the observations are so far away from the track prediction that it is more likely that all of them are false alarms, and that there was a missed detection of the target, than that one of them represents the target. The MHT filter takes this into account. Thus, even with  $K = 1$ , Sequential  $K$ -Best MHT will often outperform NN if there are missed detections. Measurement gating is effectively a crude way for NN to approximate the  $K = 1$  MHT—if



**Figure 1. Flowcharts of the NN and  $K$ -Best Trackers.**

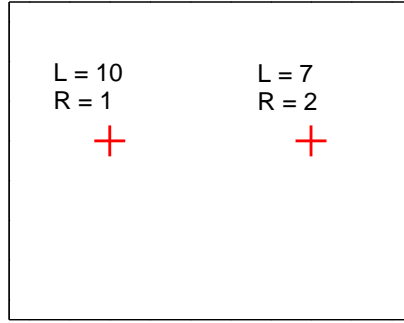
a measurement is too far away from the track, the gating causes NN to coast rather than pick the unlikely nearest neighbor.<sup>5</sup>

Sequential  $K$ -Best MHT is similar to PDA in that they are both simple recursive methods that do not require memory of past scans. For this reason, the important performance comparison of this thesis will be Sequential  $K$ -Best MHT with PDA. To our knowledge, such a comparison for these single-target trackers has not been documented.

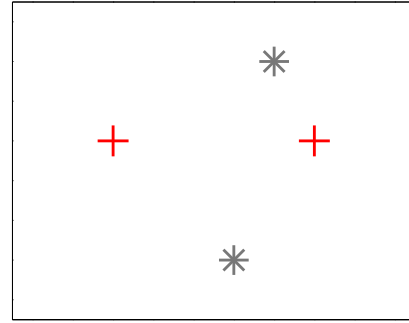
#### 1.4.2.1 Sequential $K$ -Best in the Literature

In *multiple* target tracking, a common way to keep the number of hypotheses from exploding is to maintain and propagate forward only the  $K$  best (often called the “ $m$  best”, “ $M$  best”, or “ $N$  best”) *hypotheses* from each scan, where a hypothesis is a *set* of compatible tracks. Cox and Hingorani [9] discovered that Murty’s algorithm [10] can be used to efficiently find the  $K$  best solutions to the 2-D assignment problem without using exhaustive enumeration of all of the possible hypotheses. This pruning method is a common way to

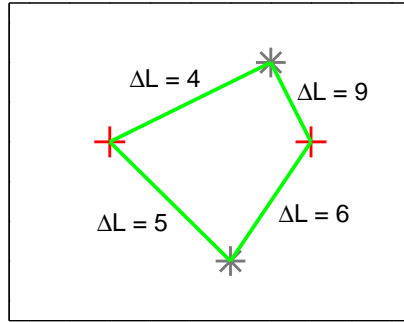
<sup>5</sup>When a “maximum likelihood gate” [8] is used, NN becomes equivalent to  $K = 1$  MHT.



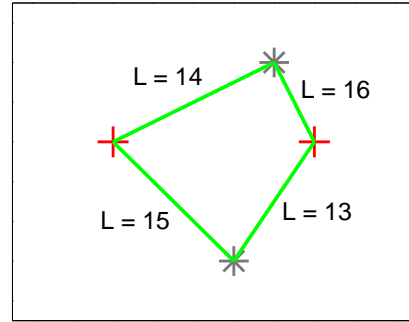
(a) Step 1



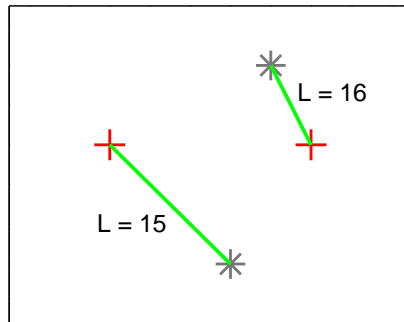
(b) Step 2



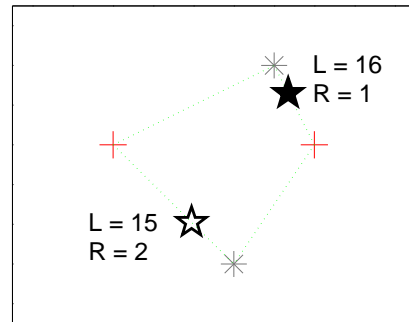
(c) Step 3



(d) Step 4



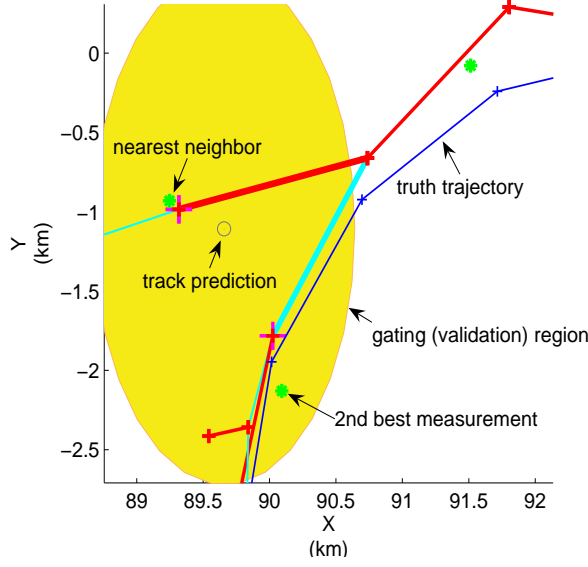
(e) Step 5



(f) Step 6

**Figure 2. Results after each of the six steps of the  $K$ -Best algorithm as numbered in the flowchart of Figure 1(b). In this example,  $K = 2$ , and we happen to also have two measurements. For simplicity, gating is not shown, and the missed-detection hypotheses are not considered (both of these steps are straightforward).  $L$  is the cumulative track score,  $\Delta L$  is the incremental score, '+' denotes a track prediction, '\*' represents a measurement, and the connecting lines are the possible data associations. The stars are filtered estimates, and the filled star is the output track state estimate of the  $K$ -Best tracker.**





**Figure 3. Illustration of the Sequential  $K$ -Best MHT tracker. Target is moving from right to left.**

implement hypothesis-oriented MHT for the multiple-target scenario. However, we have not seen the  $K$ -Best method explored in the single-target scenario, and we have not found any performance comparisons between  $K$ -Best MHT and PDA. We have also not seen the issue of track coalescence (see Chapter 4) addressed for single-target  $K$ -Best MHT.

## 1.5 Outline of the Rest of the Thesis

First, Chapter 2 explores the simulation framework used by all of our trackers. Some of the key assumptions, decisions, and implementation issues are addressed. Chapter 3 presents the simulation results of  $K$ -Best MHT and PDA, and compares the two under different tracking scenarios. Chapter 4 addresses the observed problem of track hypothesis coalescence and compares the results of various methods of detecting coalescence. Finally, a conclusion is presented.

## CHAPTER 2

### SIMULATION FRAMEWORK

We incorporated a Sequential  $K$ -Best MHT tracker into a MATLAB-based tracking framework developed primarily by Terry Ogle and Dale Blair at the Georgia Tech Research Institute. This chapter outlines the simulation framework and some of the implementation issues that were addressed.

#### 2.1 Assumptions and Filter Model

Several points need to be made about the scope of this research. All of our trackers operate in two spatial dimensions on single point targets.<sup>1</sup> Although measurements are assumed to be corrupted by noise, we neglect the effects of finite sensor resolution. We assume that binary detection decisions have already been made, and no signal-related data (i.e., signal-to-noise ratio) is used. All of these matters are important issues in real-world trackers, and will need to be considered as part of future work.

The sensor is assumed to receive range and azimuth measurements. This means the observed output  $\mathbf{y}$  is a nonlinear function of the Cartesian state  $\mathbf{x}$ , and a nonlinear estimator is needed. A standard Extended Kalman Filter (EKF) is used, which does not appear to have any problems in this application.

We use a Constant Velocity (CV) state model, which models the random target dynamics via a white noise acceleration. In our study, the model is implemented directly in the discrete-time domain and is piecewise-constant between sampling times; this differs from the discretized continuous-time model, which is also sometimes used [8].

---

<sup>1</sup>The single-target assumption is especially important, as it allows us to delve into the issues of interest without being distracted by issues such as track initiation and deletion. Extending this research to the multiple-target scenario will be the primary focus of future work.

## 2.2 Track Score Function

As mentioned in Section 1.4.2, we use the recursive probability-based score of [8] to rank the alternative association hypotheses. For the special case of detection-only data (i.e., no signal-related contributions such as SNR) in two spatial dimensions, the track score is given by

$$L(k) = L(k - 1) + \Delta L(k), \quad (1)$$

where

$$\Delta L(k) = \begin{cases} \ln[1 - P_D]; & \text{no track update on scan } k \\ \ln\left[\frac{P_D}{2\pi\beta_{FA}|\mathbf{S}|^{1/2}}\right] - \frac{d^2}{2}; & \text{track update on scan } k. \end{cases} \quad (2)$$

$P_D$  is the probability of detection and  $\beta_{FA}$  is the false alarm density. Furthermore,

$$d^2 = \tilde{\mathbf{y}}^T \mathbf{S}^{-1} \tilde{\mathbf{y}}, \quad (3)$$

where  $\tilde{\mathbf{y}}$  is the residual vector

$$\tilde{\mathbf{y}} = \mathbf{y}(k) - h(\hat{\mathbf{x}}(k|k - 1)) \quad (4)$$

and  $h(\mathbf{x})$  is the nonlinear observation function (in our case,  $h$  is the function that maps the Cartesian vector of positions and velocities to the polar vector of range and azimuth).

Finally,  $\mathbf{S}$  in (2) and (3) is defined as  $\mathbf{S} = \mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}$ , where  $\mathbf{H}$  is the linearized measurement matrix of the Extended Kalman Filter,  $\mathbf{P}$  is the prediction covariance matrix, and  $\mathbf{R}$  is the covariance matrix of the measurement noise.

Equation 2 is evaluated on each scan for both the missed-detection hypothesis (the hypothesis that there is no update on scan  $k$ ) and for each possible measurement-to-track pairing. Note that this track score function assumes knowledge of the parameters  $P_D$  and  $\beta_{FA}$ . In our research, we assume that these parameters are known and do not need to be estimated.

## 2.3 Reasonability of the MAP estimate for $K$ -Best MHT

In the explanation of  $K$ -Best MHT in Section 1.4.2, it was stated that at each scan, the track with the highest score is the output track state estimate. This section gives some brief justification of this approach.

As discussed in Section 1.3.1, one classic method to estimate the target state from the (approximate) posterior distribution seeks to minimize the squared error  $E[(\mathbf{X} - \hat{\mathbf{x}})^2]$ . The estimate  $\hat{\mathbf{x}}$  that minimizes this quadratic cost function is the mean of the posterior pdf. In  $K$ -Best MHT, a soft association approach that probabilistically combines the  $K$  estimates into one output track state estimate will thus result in lower mean-squared error than the hard decision of choosing the best track.

However, as discussed earlier, rather than minimizing the expected squared error, it may be desirable to minimize the expected “hit-or-miss” cost function (see Section 11.3 of [11]). This is equivalent to maximizing the “hit probability”

$$P_H = \int_{\hat{\mathbf{x}}-\delta}^{\hat{\mathbf{x}}+\delta} p(\mathbf{x}|\mathbf{z})d\mathbf{x} . \quad (5)$$

For “arbitrarily small”  $\delta$  (for a small “region of lethality”), the estimator that minimizes this cost function (and maximizes  $P_H$ ) is the mode of the posterior distribution, i.e., the maximum a posteriori (MAP) estimate [11].

The peak of a Gaussian mixture density depends on the means, variances, and weightings of the Gaussian components. If the means are far apart so that the “mixing” of the Gaussian components can be neglected, then the peak is determined by just the variances and the weightings. If the variances of the Gaussian components are then similar enough to be considered equal, the peak is determined by the weightings, and the MAP estimate will be the mode of the Gaussian component with the largest weighting.

In our tracking scenario, each Gaussian component represents a different track hypothesis, and the weighting (the prior probability of the track) is proportional to the track score. Therefore, the highest-scoring track can be assumed to be the MAP estimate *if* the means

of the tracks are far enough apart *and* the variances are approximately equal.

Section 4.4 will show that for Sequential  $K$ -Best MHT, there appears to be little difference in the variances of the track hypotheses. Furthermore, to remedy the problem of track coalescence, we will force the means apart from each other (see Section 4.3). Thus, the highest-scoring track will at least come close to maximizing the mixture density, and this estimate will be at least near optimal in the “hit-or-miss” sense.

## 2.4 Performance Metrics

There are many ways to define and gauge “tracking performance.” It is important to be consistent and avoid making false conclusions based on insufficient indicators of performance. When we began this research, we used two performance metrics: the mean squared error (MSE) and the track loss ratio. To calculate the mean squared error, we first ran a large number (usually between 100 and 300) of Monte Carlo runs. It is straightforward to compute a vector of the squared error for each run. Then, to get a vector of the mean squared error, one must simply average over the runs. If a track is declared “lost,” it is not used in MSE calculations, as lost tracks would make the error blow up and become meaningless. We declare a track to be lost if the track estimate,  $\hat{\mathbf{x}}(k)$ , is over 4.7 standard deviations from the true trajectory,  $\mathbf{x}(k)$ , i.e., if

$$(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T \mathbf{P}^{-1}(k)(\hat{\mathbf{x}}(k) - \mathbf{x}(k)) > 22, \quad (6)$$

for three consecutive scans.

The *track loss ratio* is the number of lost tracks in a set of Monte Carlo runs, divided by the total number of runs. It is another common tracking performance metric. The problem with trying to use both the track loss ratio and MSE metrics together is that when a track is declared lost, it is no longer used in the MSE calculations, and the average error *decreases* (since a difficult run is taken out of the average). If the track loss criteria (controlled by the ad hoc threshold) is weakened so that less tracks are lost, the result will be higher mean squared error. Therefore, when comparing two trackers, often one is better in the MSE

sense, but the other is better in the track loss ratio sense. It is impossible to say that one alternative is absolutely better than the other in such a case, so the designer must choose one metric or the other and stick with it, while realizing its shortcomings. We chose the track loss ratio as our primary metric.

Pulford [4] notes that the track loss ratio can be a misleading metric for algorithms such as PDA, which are good at not losing their true tracks, but are similarly “good” at not losing their false tracks. Thus, the track loss ratio metric might sing the praises of the PDA filter more than is fair, in comparison with the Sequential  $K$ -Best MHT.

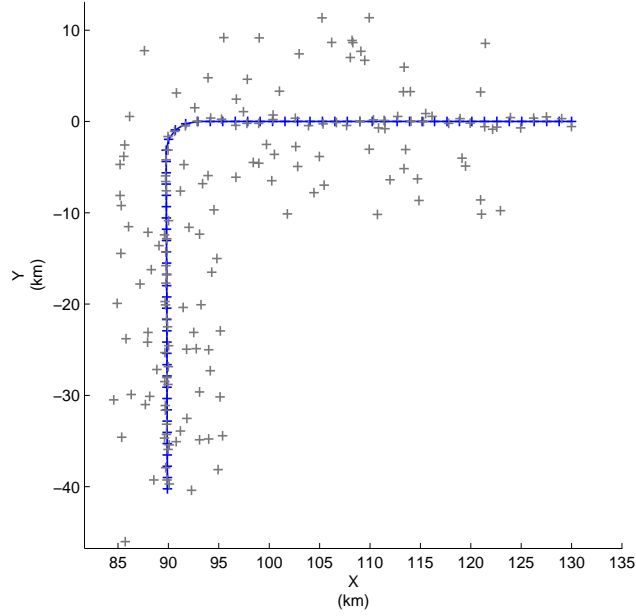
Note that in computing the track loss ratio, we look at only the single best of the  $K$  tracks (since  $\hat{\mathbf{x}}(k)$  of (6) is the single best hypothesis). Thus, it is possible that a track might be declared to be lost when the other  $K - 1$  hypotheses are still following it.

## 2.5 Truth Trajectory, False Alarms, and Gating

For all of the simulations in this thesis, the time step between measurement scans is  $T = 4$  s, and there are no time-delayed or out-of-sequence measurements. The target moves at a constant speed of 309 m/s and undergoes a 3.1 g acceleration maneuver, normal to the target velocity, for 12 s (this amounts to a 90 degree turn in the middle of the simulation). The true target trajectory and an example of the resulting measurements are shown in Figure 4. The measurements are corrupted by Gaussian noise, with range variance  $\sigma_R^2 = 1.21 \times 10^4$  m<sup>2</sup> and azimuth variance  $\sigma_\phi^2 = 0.04$  deg<sup>2</sup>; this corresponds to standard deviations of  $\sigma_R = 110$  m and  $\sigma_\phi = 0.2$  deg.

The false alarms are uniformly distributed over a “field of view” around the track. In this region, the number of false alarms per km<sup>2</sup> is Poisson-distributed with mean  $\beta_{FA}$ .

To reduce computational requirements, ellipsoidal gating is performed. Any measurements for which  $d^2 > G$  (where  $d^2$  is given in (3)) are not considered in the data association.  $G$  is set so that the gating probability is  $P_G = 0.999$ —we size the gating region so that a measurement from truth will fall within the region 99.9% of the time. This computation



**Figure 4. Truth trajectory and measurements of a simulation. Target is moving from right to left.**

of  $G$  assumes that  $d^2$  is  $\chi^2_2$ -distributed for the measurements from the truth, as explained in [8].

## 2.6 Process Noise Tuning

Under the Piecewise Constant White Acceleration (Constant Velocity) filter model, the process noise covariance matrix  $\mathbf{Q}$  is given by

$$\mathbf{Q} = \begin{bmatrix} T^4/4 & T^3/2 \\ T^3/2 & T^2 \end{bmatrix} \sigma_v^2, \quad (7)$$

where the process noise parameter  $\sigma_v^2$  is usually chosen experimentally to optimize tracking performance (see [8], [12]). Larger values of  $\sigma_v^2$  result in the Kalman Filter weighing the filter model less and the measurements more. While a larger  $\sigma_v^2$  gives the tracker more tolerance for maneuvers, it also encourages it to leave the Constant Velocity model and follow after spurious measurements. Therefore, it is important to select an appropriate process noise parameter.

Early during the course of our research, we simulated the NN and PDA trackers over a range of values for  $\sigma_v^2$ . The trackers were found to perform well with  $\sigma_v^2 = 256 \text{ (m/s}^2\text{)}^2$ . A general rule of thumb for choosing  $\sigma_v^2$  is to pick it such that

$$0.5a_m \leq \sigma_v \leq a_m, \quad (8)$$

where  $a_m$  is the maximum acceleration magnitude [8]. Since  $a_m = 3.1 \text{ g} = 30.4 \text{ m/s}^2$ , our heuristic choice of  $\sigma_v = 16 \text{ m/s}^2$  agrees with this guideline.

The question, then, is whether this same general rule for picking  $\sigma_v^2$  applies to Sequential  $K$ -Best MHT, or whether it should be modified. Due to the undesirable tendency of the  $K$  best tracks to coalescence, it is reasonable to conjecture that a higher process noise might lead to improved results.<sup>2</sup>

**Table 1. Process noise parameters tested.**

Index	$\sigma_v^2$
1	64
2	128
3	192
4	256
5	320
6	384
7	448

The  $K$ -Best tracker was simulated for each of the seven process noise parameters of Table 1. For each process noise, a table was created in the form of Table 4 (in Chapter 3). Each of these seven tables tabulated the track loss ratio over the varying probabilities of detection  $P_D$  and false alarm densities  $\beta_{FA}$ . The track loss ratio was computed over simulations of 200 Monte Carlo runs. For each combination of  $(P_D, \beta_{FA})$ , we then compared across the seven tables to find which value of  $\sigma_v^2$  resulted in the minimum track loss ratio. The results are shown in Table 2. Each entry of Table 2 is the index of the process noise parameter of Table 1 that minimized the track loss ratio for that given combination of  $(P_D, \beta_{FA})$ .

---

<sup>2</sup>We thank Paul Burns of the Georgia Tech Research Institute, who suggested that we look into this issue.



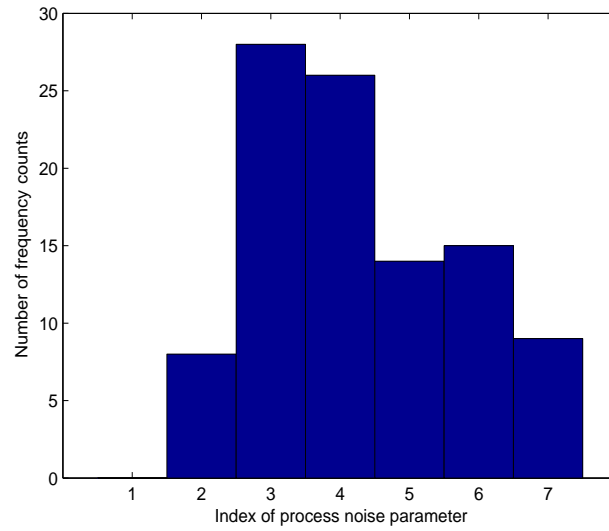
**Table 2. Which  $\sigma_v^2$  works best?**

	$P_D = 1$					$P_D = .99$					$P_D = .94$					$P_D = .87$				
$K$ :	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
$\beta_{FA} = 0$	4	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4	3	3	3	3
$\beta_{FA} = .01$	5	6	5	6	4	7	6	7	5	6	5	4	6	4	5	3	4	6	4	4
$\beta_{FA} = .04$	5	5	4	5	6	4	7	6	6	5	3	5	6	4	4	2	2	4	3	3
$\beta_{FA} = .08$	5	7	5	6	7	4	3	6	7	7	4	4	4	4	3	3	4	2	3	2
$\beta_{FA} = .13$	5	4	6	4	7	4	3	6	7	4	2	5	3	6	4	3	2	2	2	3

Under difficult tracking scenarios of high  $\beta_{FA}$  and low  $P_D$ , lower values of  $\sigma_v^2$  are desired, while higher process noise is preferred under easier scenarios. This behavior is to be expected—if there are more false alarms and missed detections, it is wise to trust the measurements less and the filter model more.

There do not appear to be any general trends relating  $K$  and  $\sigma_v^2$ . Sometimes large values of  $K$  perform best with large  $\sigma_v^2$ , while other times they prefer small  $\sigma_v^2$ .

Figure 5 is a histogram of the results of Table 2. It shows that process noises #3 and #4 are preferred in far more of the scenarios tested than the other process noises. These correspond to  $\sigma_v^2 = 192 \text{ (m/s}^2\text{)}^2$  and  $\sigma_v^2 = 256 \text{ (m/s}^2\text{)}^2$ , respectively. Since #4 is more central in the histogram, we choose  $\sigma_v^2 = 256 \text{ (m/s}^2\text{)}^2$  for all of our simulations. Thus, the heuristic (8) appears to work well in  $K$ -Best MHT.



**Figure 5. Histogram of the results presented in Table 2. The height of the graph at any of the seven process noises is the number of tracking scenarios for which that process noise was preferred over all of the others.**

## CHAPTER 3

### COMPARING $K$ -BEST MHT WITH PDA

The first major thrust of this research is to determine under what general conditions Sequential  $K$ -Best data association is preferable to PDA. Both data association techniques were simulated with varying probability of detection  $P_D$  and false alarm density  $\beta_{FA}$ .

Tables 3 and 4 contain the track loss ratios for a PDA and a Sequential  $K$ -Best MHT tracker, respectively. All the simulations in this thesis use 200 Monte Carlo runs. The  $K$ -Best MHT algorithm in Table 4 uses a measurement constraint (explained in Section 4.1) to prevent track coalescence. In Table 4, the track loss ratio is seen to generally decrease with increasing  $K$ , as expected.

**Table 3. Track loss ratio of PDA.**

	$P_D = 1$	$P_D = .99$	$P_D = .94$	$P_D = .87$
$\beta_{FA} = 0$	.00	.00	.00	.00
$\beta_{FA} = .01$	.04	.02	.06	.12
$\beta_{FA} = .04$	.11	.14	.27	.46
$\beta_{FA} = .08$	.18	.37	.54	.78
$\beta_{FA} = .13$	.48	.55	.79	.93

**Table 4. Track loss ratio of  $K$ -Best MHT (with measurement constraint to prevent track coalescence).**

$K$ :	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.01	.00	.00	.11	.02	.00	.00	.21	.02	.00	.01	.41	.06	.09	.03
$\beta_{FA} = .04$	.27	.04	.04	.01	.38	.08	.05	.01	.60	.24	.19	.07	.77	.58	.46	.39
$\beta_{FA} = .08$	.54	.11	.10	.10	.67	.31	.12	.17	.84	.51	.50	.56	.95	.89	.79	.82
$\beta_{FA} = .13$	.79	.38	.23	.27	.82	.45	.40	.41	.94	.80	.77	.78	1.00	.97	.95	.96

There are some places where the track loss ratio appears to actually *increase* with  $K$ ; for example, when  $\beta_{FA} = .08$ ,  $K = 3$  appears to be better than  $K = 4$ . However, as shown in Table 5, when 10,000 Monte Carlo runs were used on a problematic section of this table, all of the “hiccups” in the data vanished and performance consistently decreased and levelled

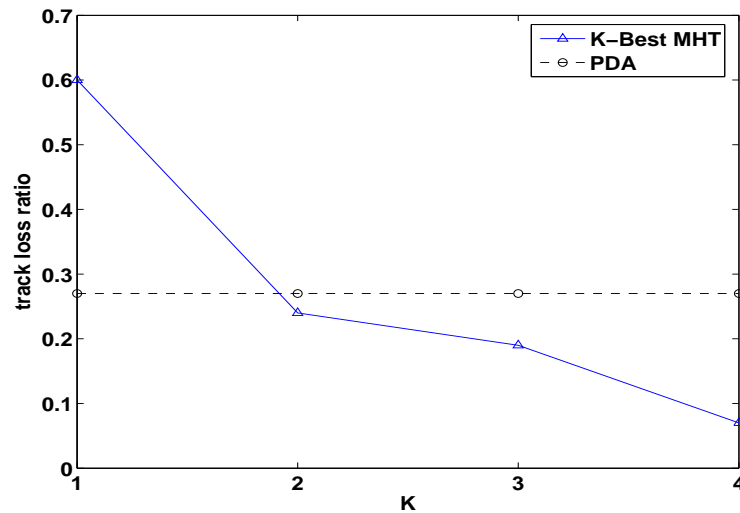
off with  $K$ . Thus, this phenomenon appears to merely be an artifact of a modest number of Monte Carlo runs.

**Table 5. The bottom left corner of Table 4, but now with 10,000 Monte Carlo Runs.**

	$P_D = 1$				$P_D = .99$				$P_D = .94$			
$K$ :	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = .04$	.29	.04	.02	.02	.39	.07	.04	.04	.60	.25	.17	.16
$\beta_{FA} = .08$	.53	.15	.09	.09	.66	.25	.17	.16	.85	.58	.52	.50
$\beta_{FA} = .13$	.74	.37	.27	.24	.85	.49	.40	.39	.94	.81	.79	.78

Simulations were also run for greater values of  $K$ , but when  $K > 4$ , the track loss ratios level off and give diminishing returns. This same effect was observed when 10,000 Monte Carlo runs were used.

Tables 3 and 4 indicate that under difficult tracking scenarios of high false alarm density and low probability of detection, PDA is comparable to Sequential  $K$ -Best MHT, and neither is particularly advantageous over the other. Otherwise, and in most cases,  $K$ -Best association generally outperforms PDA if  $K \geq 2$ . Figure 6 illustrates this by plotting the track loss ratio as a function of  $K$  for the case in which  $P_D = .94$ ,  $\beta_{FA} = .04$ .

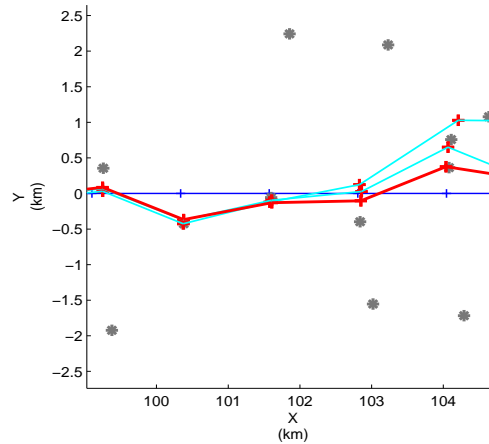


**Figure 6. Track loss ratio for the  $P_D = 0.94, \beta_{FA} = 0.04$  case.**

## CHAPTER 4

### DEALING WITH TRACK COALESCENCE

Upon implementing the Sequential  $K$ -Best MHT tracker, we observed a fundamental problem: the  $K$  best track hypotheses would often coalesce and become effectively one track. If two track hypotheses are similar or close to each other, they are likely to pick the same measurement during data association. Updating the tracks with the same measurement generally makes them more similar to each other, making this more likely to happen again on the subsequent scan. Track hypotheses quickly coalesce to the point that we have, in effect, just a single-hypothesis tracker (see Figure 7). The second major thrust of this thesis is to compare different approaches to resolve this issue.



**Figure 7. Illustration of track coalescence when  $K = 3$ . Target is moving from right to left.**

#### 4.1 The Simple “Measurement Constraint” Method

We formulated a number of methods to prevent track coalescence. The first approach, which we call the “Measurement Constraint” method, simply prevents different tracks from picking the same measurement during data association. Table 4, which was analyzed earlier, uses this “measurement constraint.” Table 6 contains the results for when nothing is done to prevent track coalescence.

**Table 6. Track loss ratio of  $K$ -Best MHT without any method to prevent track coalescence; compare with Table 4.**

	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
$K$ :	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.02	.01	.01	.11	.02	.01	.01	.21	.07	.04	.02	.41	.14	.14	.09
$\beta_{FA} = .04$	.27	.11	.03	.08	.38	.23	.14	.09	.60	.43	.33	.27	.77	.69	.56	.60
$\beta_{FA} = .08$	.54	.32	.20	.26	.67	.44	.31	.24	.84	.72	.54	.60	.95	.89	.86	.87
$\beta_{FA} = .13$	.79	.52	.48	.36	.82	.66	.53	.54	.94	.88	.84	.88	1.00	.98	.97	.98

For almost all of the values of  $P_D$ ,  $\beta_{FA}$ , and  $K$ , use of the measurement constraint method results in a lower track loss ratio than when no method is used to prevent track coalescence (except for when  $K = 1$ , in which case the two methods perform identically, as expected). Thus, this crude measurement constraint technique inspires research into other, more rigorous, methods.

## 4.2 Tighter Coalescence Tests

The obvious problem with the measurement constraint method is that two track hypotheses can be different from each other and yet rightfully want to pick the same measurement. For this reason, it seems necessary to test the similarity of the track state vectors before forcing them to pick different measurements. In other words, a *tighter* test to determine track coalescence might be desired, one that will not declare tracks to be coalesced so easily. This section explores two such tests.

### 4.2.1 Tighter Test I

In the measurement constraint approach described above, two track hypotheses are declared similar if they pick the same measurement. A logical extension of this idea would be to use an “ $M$  of  $N$ ” approach: the tracks are declared to be coalesced if they pick the same measurements in  $M$  out of the last  $N$  scans (if  $M = N$ , this reduces to the requirement that they pick the same  $M$  consecutive measurements). This is a tighter test than merely requiring the tracks to pick the same measurement on a single scan; however, it requires a history of data associations. One of the attractive characteristic features of the Sequential

$K$ -Best algorithm is that it does not require any such history to be maintained. This makes the “ $M$  of  $N$ ” approach unsatisfactory.

We explored a variation of the  $M$  of  $N$  approach that does not require any history.<sup>1</sup> In this variation, each track hypothesis has an integer score associated with it that is incremented every time the track picks the same observation as another track. If the track picks a measurement that no other track has picked, the score is decremented. The score is not allowed to go below zero. If the score reaches the predetermined threshold, the track hypothesis is declared to be coalesced with another. For example, if the threshold is set to 3, then this method will declare a track to be similar to another if it shares 3 of 3, or 4 of 5, or 5 of 7 (etc.) measurements with other tracks. (One problem with this method is that if a track is known to have shared measurements with another track over the last few scans, there is no way of knowing whether it was with a single other track, or whether it was with different tracks on different scans.) Once the threshold is reached, the measurement constraint comes into play and the track is not allowed to pick a measurement that has already been picked by another track.

This test was used in simulations over the same range of values of  $P_D$ ,  $\beta_{FA}$ , and  $K$  as Table 6. The full results with thresholds of 2, 3, and 4 are tabulated in Appendix A. The three thresholds generally led to similar performance over the scenarios tested. Surprisingly, this method appears to overall be *worse* than the simple measurement constraint technique under the scenarios tested. Figure 8 shows one particular case of  $P_D = .94$ ,  $\beta_{FA} = .04$ .

#### 4.2.2 Tighter Test II

The approaches to recognizing track coalescence described above have been based on the measurements—if two tracks pick the same measurements, they are assumed to be similar. We now explore a different method, the “Mahalanobis Distance,” which has been used to test track similarity in the literature<sup>2</sup> (for example, see [13]). The Mahalanobis distance  $d_1$

<sup>1</sup>We wish to thank Dale Blair of the Georgia Tech Research Institute, for suggesting this idea to us.

<sup>2</sup>The motive to identify similar tracks is different for most MHT methods than for Sequential  $K$ -Best MHT. Whereas most MHT methods identify similar tracks in order to merge them and make the algorithm

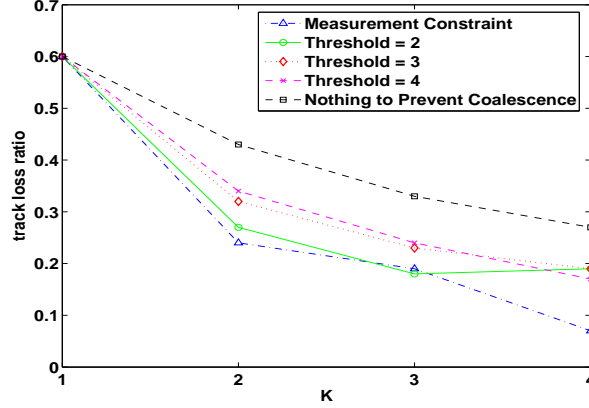


Figure 8. Track loss ratio when  $P_D = .94, \beta_{FA} = .04$ .

between two tracks  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is given by

$$d_1^2 = (\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{P}_B^{-1} (\mathbf{x}_1 - \mathbf{x}_2), \quad (9)$$

where  $\mathbf{P}_B$  is the covariance matrix of the higher-scoring of the two tracks. The tracks are then declared similar if  $d_1 < \lambda$ , where  $\lambda$  is a predetermined threshold (chosen experimentally).

The measurement constraint is used together with this statistical distance test. If two tracks<sup>3</sup> are declared to be similar by the Mahalanobis distance test, then they are not allowed to pick the same measurement.

This method was tested over a range of values of  $P_D$ ,  $\beta_{FA}$ , and  $K$ , and the threshold  $\lambda$  was also varied. Once again, this method was consistently *worse* than the simple measurement constraint technique under the scenarios tested. Figure 9 shows a typical scenario of  $P_D = .94, \beta_{FA} = .04$ . The bottom curve is for  $\lambda = \infty$ , which is the same thing as the measurement constraint method without any similarity test. The top curve is for  $\lambda = 0$ , which corresponds to no method being used to prevent track coalescence. The middle curve is for  $\lambda = 0.5$ , which is better than when nothing is done to prevent track coalescence ( $\lambda = 0$ ),

computationally feasible,  $K$ -Best MHT is already computationally feasible (due to the  $K$ -Best constraint) and needs to identify similar tracks to prevent the track coalescence from robbing us of the ability to keep track of other alternatives. Whatever the motivation, however, the issue of *recognizing* track similarity should be the same in both cases.

<sup>3</sup>An alternative is to measure the Mahalanobis distance between the track predictions  $\hat{\mathbf{x}}(k|k-1)$ . This led to very similar results.



but worse than the original measurement constraint ( $\lambda = \infty$ ).

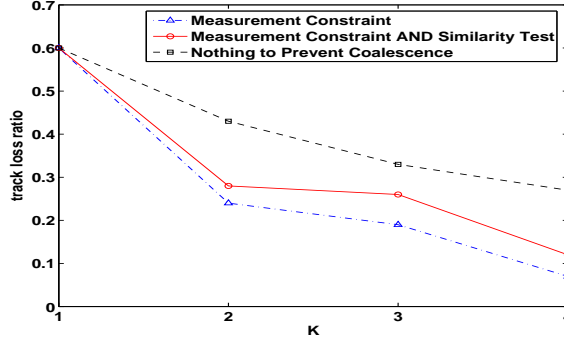


Figure 9. Track loss ratio when  $P_D = .94, \beta_{FA} = .04$ .

Figure 9 represents the clear trend in our simulations: increasing  $\lambda$ , i.e., decreasing the influence of the similarity test, consistently leads to better tracker performance. This observation, together with the results of the previous section, suggest an important conclusion: *declaring tracks to be coalesced when they are actually not may be a problem, but it is apparently not as detrimental as letting track coalescence go undetected*. These results persuade us to change our earlier course and pursue “loose” criteria for detecting track coalescence.

### 4.3 Looser Coalescence Tests

One simple way to detect track coalescence is to use the Mahalanobis distance test of (9) by itself, without the measurement-picking constraint. However, once tracks are declared to be coalesced, what should be done about it? We tried two basic remedies. The first remedy adds Gaussian noise to the worse of the two tracks: If  $\mathbf{x}_2$  is the track with the lower score, with covariance  $\mathbf{P}_2$ , then let

$$\mathbf{x}_2^* = \mathbf{x}_2 + \alpha(\mathbf{P}_2)^{1/2}\mathbf{v}, \quad (10)$$

where  $(\mathbf{P}_2)^{1/2}$  is the matrix square root of  $\mathbf{P}_2$ ,  $\mathbf{v} = (v_1, \dots, v_N)$ , where each  $v_i$  is independently distributed according to a standard normal distribution, and  $\alpha$  is an experimentally chosen scalar that determines how much random perturbation is added to the track.

A possible problem with this idea is that it is just as likely to bring the tracks closer together as it is to push them apart. A variation of this is to push the lower-scoring track away from the higher-scoring one:

$$\mathbf{x}_2^* = \mathbf{x}_2 + \alpha(\mathbf{x}_2 - \mathbf{x}_1). \quad (11)$$

In most of our simulations, the first remedy described above led to slightly better performance than the second, so in all of our similarity tests, we use (10) as a remedy once the tracks are declared to be coalesced. Note that if two track hypotheses are declared similar by the Mahalanobis distance test, we *always* modify the lower-scoring hypothesis, whereas in the previous section, it might be the case that two tracks naturally pick different measurements and the measurement constraint does not come into play. Thus, this is a “looser” test than in the previous section.

We tried the Mahalanobis similarity test along with two variations of it. In (9),  $\mathbf{P}_B$  was the covariance matrix of the better of the two tracks. An alternative that is often used (for example, see [13]) is to let  $\mathbf{P}$  be the sum of the covariance matrices of the two tracks. In this case, the new distance is

$$d_2^2 = (\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{P}_1 + \mathbf{P}_2)^{-1} (\mathbf{x}_1 - \mathbf{x}_2). \quad (12)$$

A second alternative is to average the Mahalanobis distances (or rather, the square of the distances) that result from using  $\mathbf{P}_1$  and  $\mathbf{P}_2$ . In this case, the distance is given by

$$d_3^2 = \frac{(\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}) (\mathbf{x}_1 - \mathbf{x}_2)}{2}. \quad (13)$$

These three variations of the statistical distance test to detect track coalescence were used in simulations over the same range of values of  $P_D$ ,  $\beta_{FA}$ , and  $K$  as Table 6, and the threshold  $\lambda$  was also varied. The full results are tabulated in Appendix B. The three methods lead to similar tracking performance, and over all of the simulations, none of the three is consistently better or worse than the others. Figure 10 shows a typical scenario of  $P_D = .94$ ,

$\beta_{FA} = .04$ . All three are consistently better than when nothing is done to prevent track coalescence. However, all three are also consistently *worse* than the simple measurement constraint technique.

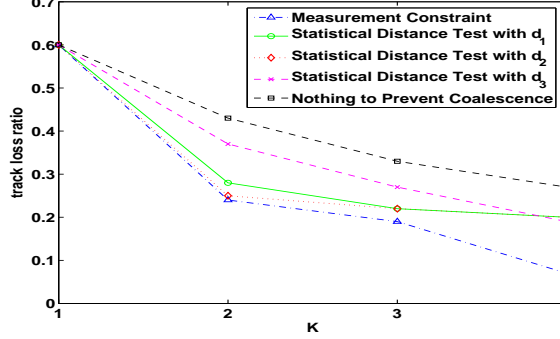


Figure 10. Track loss ratio when  $P_D = .94, \beta_{FA} = .04$ .

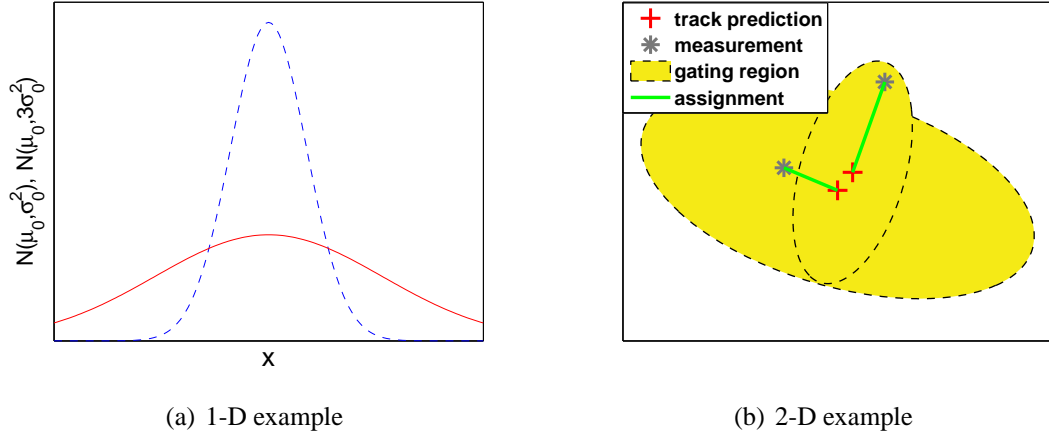
#### 4.4 The J-Divergence Test

The basic shortcoming of the track similarity test described in the previous section is that two track hypotheses might have similar track state vectors, yet have quite different covariances associated with them. Such track hypotheses are not truly similar and may not be as likely to cause coalescence problems (see Figure 11). Reid points out the importance of combining tracks that have “similar effects,” and explains that two tracks have similar effects if the means *and* variances are “sufficiently close” [7]. However, he does not explain *how* he determines if the means and variances are sufficiently close. We use the so-called J-Divergence, which is a symmetrized Kullback-Leibler distance, to make this decision.

The Kullback-Leibler (KL) distance is a natural measure of the dissimilarity between two probability distributions. If  $\mathbf{X}_1 \sim p(x_1, \dots, x_N)$  and  $\mathbf{X}_2 \sim q(x_1, \dots, x_N)$ , then the Kullback-Leibler distance  $\mathcal{D}(p||q)$  is defined as

$$\mathcal{D}(p||q) = \int p(x_1, \dots, x_N) \ln \frac{p(x_1, \dots, x_N)}{q(x_1, \dots, x_N)} dx_1 \dots dx_N. \quad (14)$$

If track hypotheses  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are treated as  $N$ -variate random vectors  $\mathbf{X}_1 \sim p(x_1, \dots, x_N)$



**Figure 11.** Even if two tracks are close together, they are likely to pick different measurements if the covariances are different.

and  $\mathbf{X}_2 \sim q(x_1, \dots, x_N)$ ,<sup>4</sup>  $\mathcal{D}(p||q)$  will measure the overall dissimilarity in the distributions. This is more theoretically appealing than the Mahalanobis distance, which measures only the difference of the means of the track state distributions.

In Appendix C, the KL distance from one  $N$ -variate Gaussian distribution,  $\mathcal{N}_2(\mathbf{x}_2, \mathbf{P}_2)$ , to another,  $\mathcal{N}_1(\mathbf{x}_1, \mathbf{P}_1)$ , is shown to be given by

$$\mathcal{D}(\mathcal{N}_2||\mathcal{N}_1) = \frac{1}{2} [(\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{P}_1^{-1} (\mathbf{x}_1 - \mathbf{x}_2)] \quad (15a)$$

$$+ \frac{1}{2} [\text{tr}[\mathbf{P}_1^{-1} \mathbf{P}_2] - \ln |\mathbf{P}_2 \mathbf{P}_1^{-1}| - N]. \quad (15b)$$

Equation (15) averages two terms; the first, (15a), is the Mahalanobis distance used earlier to determine the difference between the means of the two random vectors, and the second, (15b), is a new term that measures the difference in the covariances.

The KL distance is asymmetric, i.e.,  $\mathcal{D}(p||q) \neq \mathcal{D}(q||p)$ . For example, Figure 11(b) shows two Gaussians with  $\mu_1 = \mu_2$ ,  $\sigma_2 = 3\sigma_1$ . For this scalar case, (15) gives

$$\begin{aligned} \mathcal{D}(p||q) &= 4 - \ln 3 = 2.9 \\ \mathcal{D}(q||p) &= \ln 3 - \frac{4}{9} = 0.65. \end{aligned} \quad (16)$$

<sup>4</sup>The notation may be confusing—when the literature refers to a “track” as  $\mathbf{x}_i$ , it is usually referring to the *track state estimate*. Here we treat the track hypothesis as a random vector  $\mathbf{X}_i$ , and  $\mathbf{x}_i$  is the mean of  $\mathbf{X}_i$ .

Since  $\mathcal{D}(p\|q) \gg \mathcal{D}(q\|p)$ , the “distance” from  $\mathbf{X}_1$  to  $\mathbf{X}_2$  is much different than the “distance” from  $\mathbf{X}_2$  to  $\mathbf{X}_1$ , so the KL distance, by itself, may not be a good measure of the “distance between” the two tracks.

To deal with this, we simply average the two possible KL distances to yield the so-called J-Divergence:

$$\mathcal{J}(p\|q) = \frac{\mathcal{D}(p\|q) + \mathcal{D}(q\|p)}{2}. \quad (17)$$

The J-Divergence is attractive because it is symmetric, i.e.,  $\mathcal{J}(p\|q) = \mathcal{J}(q\|p)$ . Applying (17) to (15) gives

$$\begin{aligned} \mathcal{J}(\mathcal{N}_1\|\mathcal{N}_2) &= \frac{1}{2} [\mathcal{D}(\mathcal{N}_2\|\mathcal{N}_1) + \mathcal{D}(\mathcal{N}_1\|\mathcal{N}_2)] \\ &= \frac{1}{2} \left\{ \frac{1}{2} (\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}) (\mathbf{x}_1 - \mathbf{x}_2) \right. \\ &\quad \left. + \frac{1}{2} [\text{tr}[\mathbf{P}_1^{-1}\mathbf{P}_2] + \text{tr}[\mathbf{P}_2^{-1}\mathbf{P}_1] - (\ln|\mathbf{P}_2\mathbf{P}_1^{-1}| + \ln|\mathbf{P}_1\mathbf{P}_2^{-1}|) - 2N] \right\}. \quad (18) \end{aligned}$$

Working with the second term of (18),

$$\text{tr}[\mathbf{P}_1^{-1}\mathbf{P}_2] + \text{tr}[\mathbf{P}_2^{-1}\mathbf{P}_1] = \text{tr}[\mathbf{P}_1^{-1}\mathbf{P}_2 + \mathbf{P}_2^{-1}\mathbf{P}_1], \quad \text{and} \quad (19)$$

$$\ln|\mathbf{P}_2\mathbf{P}_1^{-1}| + \ln|\mathbf{P}_1\mathbf{P}_2^{-1}| = \ln|\mathbf{P}_2\mathbf{P}_1^{-1}||\mathbf{P}_1\mathbf{P}_2^{-1}| = \ln|\mathbf{P}_2\mathbf{P}_1^{-1}\mathbf{P}_1\mathbf{P}_2^{-1}| = \ln|I| = \ln(1) = 0. \quad (20)$$

Therefore,

$$\mathcal{J}(\mathcal{N}_1\|\mathcal{N}_2) = \underbrace{\frac{1}{2} \left[ \frac{(\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}) (\mathbf{x}_1 - \mathbf{x}_2)}{2} \right]}_{\mathcal{J}_{mean}} + \underbrace{\frac{1}{2} \left[ \frac{\text{tr}[\mathbf{P}_2^{-1}\mathbf{P}_1 + \mathbf{P}_1^{-1}\mathbf{P}_2] - 2N}{2} \right]}_{\mathcal{J}_{var}}. \quad (21)$$

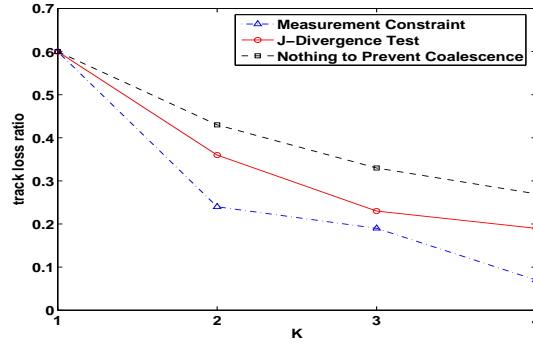
Equation (21) is also the average of two terms; the first,  $\mathcal{J}_{mean}$ , is the average of the two possible Mahalanobis distances and is equal to  $d_3^2$  of (13), while the second,  $\mathcal{J}_{var}$ , is the average of the terms due to the variance. This decomposition of the J-Divergence will be useful in later analysis.

Using our J-Divergence method, the tracks  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , with distributions  $\mathcal{N}_1(\mathbf{x}_1, \mathbf{P}_1)$  and  $\mathcal{N}_2(\mathbf{x}_2, \mathbf{P}_2)$ , respectively, are declared to be coalesced if  $\mathcal{J}(\mathcal{N}_1\|\mathcal{N}_2) < \lambda$ .

The results of the simulations that used the J-Divergence test are shown in Table 7. These results are similar to those of the Mahalanobis distance test (see Figure 12 for one scenario of  $\beta_{FA} = .04$ ,  $P_D = .94$ ).

**Table 7. Track loss ratio of  $K$ -Best MHT with the J-Divergence Test.**

$K$ :	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.02	.00	.00	.11	.01	.00	.00	.21	.05	.03	.02	.41	.14	.09	.08
$\beta_{FA} = .04$	.27	.04	.04	.02	.38	.13	.02	.06	.60	.36	.23	.19	.77	.68	.57	.48
$\beta_{FA} = .08$	.54	.24	.11	.08	.67	.27	.17	.17	.84	.66	.56	.56	.95	.87	.85	.83
$\beta_{FA} = .13$	.79	.44	.31	.27	.82	.59	.47	.45	.94	.84	.79	.83	1.00	.99	.97	.95



**Figure 12. Track loss ratio when  $P_D = .94$ ,  $\beta_{FA} = .04$ .**

The J-divergence test did not bring much improvement over the Mahalanobis test because it turned out to be practically the same thing; over a simulation of 200 Monte Carlo runs,  $\mathcal{J}_{mean}$  of (21) accounted for 99.0% of the total  $\mathcal{J}$  on average, completely dwarfing the other term,  $\mathcal{J}_{var}$ . The means of the track state distributions differed from each other much more than the variances, and the J-Divergence test was essentially equivalent to the statistical distance test in our Sequential  $K$ -Best Multiple Hypothesis Tracking experiments.

## CHAPTER 5

### CONCLUSION

Several conclusions may be drawn from our simulations of single-target trackers. Under difficult tracking scenarios of high false alarm density and low probability of detection, *K*-Best MHT gives results comparable to PDA. Otherwise, in most cases, the *K*-Best association generally begins to outperform the PDA if  $K \geq 2$ .

Several methods were tested to detect track hypothesis coalescence. All of them led to better tracking performance than when nothing was done to prevent coalescence. The best results were generally found with the “looser” tests, i.e., those that more readily declared tracks to be coalesced. Using the J-Divergence to measure the dissimilarity between tracks led to essentially the same results as using the Mahalanobis distance and its variations, since the difference in means dominated the J-Divergence. Surprisingly, the most effective method to deal with track coalescence was to simply not let two track hypotheses pick the same measurement during data association.

Future work will focus on extending this research into the more general multiple-target scenario. The main purpose for focusing on the single-target case in this thesis was to gain insight that could later be applied to the more complex multiple-target case.

## APPENDIX A

### RESULTS OF “TIGHTER TEST I” OF SECTION 4.2.1

**Table 8. Track loss ratio of  $K$ -Best MHT when the threshold is 2.**

	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
$K$ :	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.02	.00	.01	.11	.02	.01	.00	.21	.03	.02	.01	.41	.12	.06	.03
$\beta_{FA} = .04$	.27	.06	.02	.02	.38	.10	.06	.04	.60	.27	.18	.19	.77	.56	.48	.46
$\beta_{FA} = .08$	.54	.20	.15	.10	.67	.28	.18	.20	.84	.60	.54	.51	.95	.84	.78	.84
$\beta_{FA} = .13$	.79	.45	.34	.28	.82	.56	.41	.43	.94	.81	.73	.76	1.00	.98	.97	.97

**Table 9. Track loss ratio of  $K$ -Best MHT when the threshold is 3.**

	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
$K$ :	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.02	.00	.01	.11	.01	.00	.01	.21	.03	.01	.03	.41	.13	.07	.07
$\beta_{FA} = .04$	.27	.05	.04	.03	.38	.08	.04	.04	.60	.32	.23	.19	.77	.60	.41	.44
$\beta_{FA} = .08$	.54	.24	.08	.09	.67	.30	.20	.17	.84	.57	.56	.52	.95	.82	.86	.80
$\beta_{FA} = .13$	.79	.44	.27	.28	.82	.54	.46	.44	.94	.87	.76	.85	1.00	.97	.95	.96

**Table 10. Track loss ratio of  $K$ -Best MHT when the threshold is 4.**

	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
$K$ :	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.01	.00	.00	.11	.01	.01	.01	.21	.03	.02	.01	.41	.11	.04	.05
$\beta_{FA} = .04$	.27	.06	.04	.02	.38	.10	.05	.06	.60	.34	.24	.17	.77	.63	.51	.40
$\beta_{FA} = .08$	.54	.24	.10	.10	.67	.32	.14	.21	.84	.54	.55	.58	.95	.91	.79	.88
$\beta_{FA} = .13$	.79	.42	.33	.31	.82	.51	.50	.46	.94	.81	.81	.76	1.00	.97	.95	.96



## APPENDIX B

### RESULTS OF THE MAHALANOBIS TESTS OF SECTION 4.3

**Table 11. Track loss ratio of  $K$ -Best MHT when  $d_1^2$  is used as the Mahalanobis distance.**

$K$ :	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.02	.01	.01	.11	.02	.01	.01	.21	.04	.01	.03	.41	.17	.09	.07
$\beta_{FA} = .04$	.27	.07	.04	.01	.38	.10	.07	.04	.60	.28	.22	.20	.77	.64	.47	.48
$\beta_{FA} = .08$	.54	.24	.11	.07	.67	.22	.25	.20	.84	.63	.60	.59	.95	.91	.87	.84
$\beta_{FA} = .13$	.79	.44	.33	.30	.82	.57	.41	.44	.94	.88	.85	.80	1.00	.94	.95	.96

**Table 12. Track loss ratio of  $K$ -Best MHT when  $d_2^2$  is used as the Mahalanobis distance.**

$K$ :	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.01	.00	.00	.11	.02	.00	.01	.21	.03	.03	.03	.41	.11	.09	.07
$\beta_{FA} = .04$	.27	.06	.03	.03	.38	.14	.07	.06	.60	.25	.22	.20	.77	.63	.50	.56
$\beta_{FA} = .08$	.54	.21	.11	.13	.67	.35	.24	.14	.84	.66	.60	.51	.95	.88	.87	.88
$\beta_{FA} = .13$	.79	.42	.30	.23	.82	.56	.48	.40	.94	.82	.78	.79	1.00	.96	.95	.95

**Table 13. Track loss ratio of  $K$ -Best MHT when  $d_3^2$  is used as the Mahalanobis distance.**

$K$ :	$P_D = 1$				$P_D = .99$				$P_D = .94$				$P_D = .87$			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
$\beta_{FA} = 0$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$\beta_{FA} = .01$	.07	.00	.01	.01	.11	.03	.01	.01	.21	.03	.01	.03	.41	.14	.11	.10
$\beta_{FA} = .04$	.27	.04	.02	.02	.38	.08	.07	.05	.60	.37	.27	.19	.77	.63	.50	.52
$\beta_{FA} = .08$	.54	.23	.12	.11	.67	.30	.23	.17	.84	.69	.54	.61	.95	.91	.87	.89
$\beta_{FA} = .13$	.79	.40	.28	.34	.82	.59	.46	.45	.94	.86	.80	.79	1.00	.99	.96	.96

## APPENDIX C

### KULLBACK-LEIBLER DISTANCE FOR MULTIVARIATE GAUSSIAN NORMAL DISTRIBUTIONS

This appendix derives the Kullback-Leibler (KL) distance from one Gaussian vector to another.<sup>1</sup> As stated in (14), if  $\mathbf{X}_1 \sim p(x_1, \dots, x_N) = p(\mathbf{x})$  and  $\mathbf{X}_2 \sim q(x_1, \dots, x_N) = q(\mathbf{x})$ , then the KL distance  $\mathcal{D}(p||q)$  is defined<sup>2</sup> as

$$\begin{aligned}\mathcal{D}(p||q) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, \dots, x_N) \ln \frac{p(x_1, \dots, x_N)}{q(x_1, \dots, x_N)} dx_1 \dots dx_N \\ &= \int_{\mathbf{R}^N} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}.\end{aligned}\tag{22}$$

If  $\mathbf{X}_1 \sim \mathcal{N}_1(\mathbf{x}_1, \mathbf{P}_1)$  and  $\mathbf{X}_2 \sim \mathcal{N}_2(\mathbf{x}_2, \mathbf{P}_2)$ , then

$$\begin{aligned}p(\mathbf{x}) &= \frac{1}{(2\pi)^{N/2} |\mathbf{P}_2|^{1/2}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2) \right], \\ q(\mathbf{x}) &= \frac{1}{(2\pi)^{N/2} |\mathbf{P}_1|^{1/2}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_1)^T \mathbf{P}_1^{-1}(\mathbf{x} - \mathbf{x}_1) \right],\end{aligned}$$

and therefore

$$\ln \frac{p(\mathbf{x})}{q(\mathbf{x})} = \ln \frac{|\mathbf{P}_1|^{1/2}}{|\mathbf{P}_2|^{1/2}} + \frac{1}{2}(\mathbf{x} - \mathbf{x}_1)^T \mathbf{P}_1^{-1}(\mathbf{x} - \mathbf{x}_1) - \frac{1}{2}(\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2). \tag{23}$$

Plugging (23) into (22),

$$\begin{aligned}\mathcal{D}(\mathcal{N}_2||\mathcal{N}_1) = \mathcal{D}(p||q) &= \int p(\mathbf{x}) \ln \frac{|\mathbf{P}_1|^{1/2}}{|\mathbf{P}_2|^{1/2}} d\mathbf{x} + \frac{1}{2} \int p(\mathbf{x}) (\mathbf{x} - \mathbf{x}_1)^T \mathbf{P}_1^{-1}(\mathbf{x} - \mathbf{x}_1) d\mathbf{x} \\ &\quad - \frac{1}{2} \int p(\mathbf{x}) (\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2) d\mathbf{x} \\ &= \ln \frac{|\mathbf{P}_1|^{1/2}}{|\mathbf{P}_2|^{1/2}} \int p(\mathbf{x}) d\mathbf{x} \\ &\quad + \frac{1}{2} \int \frac{1}{(2\pi)^{N/2} |\mathbf{P}_2|^{1/2}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2) \right] (\mathbf{x} - \mathbf{x}_1)^T \mathbf{P}_1^{-1}(\mathbf{x} - \mathbf{x}_1) d\mathbf{x} \\ &\quad - \frac{1}{2} \int \frac{1}{(2\pi)^{N/2} |\mathbf{P}_2|^{1/2}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2) \right] (\mathbf{x} - \mathbf{x}_2)^T \mathbf{P}_2^{-1}(\mathbf{x} - \mathbf{x}_2) d\mathbf{x}\end{aligned}$$

<sup>1</sup>Kullback himself gives some related results (but without complete derivation) in [14]. A special case of this KL distance is also used for a quite different application in [15].

<sup>2</sup>Since  $\int p \log_2(p/q) = \int p(\log_2 e) \ln(p/q) = (\log_2 e) \int p \ln(p/q)$ , this definition differs from the common  $\log_2$ -based definition by only a scale factor.

$$\begin{aligned}
&= \frac{1}{2} \ln \frac{|\mathbf{P}_1|}{|\mathbf{P}_2|} + \frac{1}{2} \int \frac{1}{(2\pi)^{N/2} |\mathbf{P}_2|^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} \right] (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) d\mathbf{s} \\
&\quad - \frac{1}{2} \int \frac{1}{(2\pi)^{N/2} |\mathbf{P}_2|^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} \right] \mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} d\mathbf{s} \\
&= -\frac{1}{2} \ln \frac{|\mathbf{P}_2|}{|\mathbf{P}_1|} + \frac{1}{2} \int f(\mathbf{s}) (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) d\mathbf{s} - \frac{1}{2} \int f(\mathbf{s}) \mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} d\mathbf{s}, \quad (27)
\end{aligned}$$

where a change of variables,  $\mathbf{s} = \mathbf{x} - \mathbf{x}_2$ , was made to write the expression in terms of the zero-mean multivariate normal density  $f$  of the random vector  $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_2)$ .

The quadratic forms  $(\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)$  and  $\mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s}$  each evaluate to a scalar.

Thus,

$$(\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) = \text{tr} \left[ (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) \right] \quad (28)$$

and

$$\mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} = \text{tr} \left[ \mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} \right]. \quad (29)$$

Combining this observation with the cyclic property of trace ( $\text{tr}[\mathbf{ABC}] = \text{tr}[\mathbf{CAB}] = \text{tr}[\mathbf{BCA}]$ ), we have

$$(\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) = \text{tr} \left[ \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \right] \quad (30)$$

and

$$\mathbf{s}^T \mathbf{P}_2^{-1} \mathbf{s} = \text{tr} \left[ \mathbf{P}_2^{-1} \mathbf{s} \mathbf{s}^T \right]. \quad (31)$$

Therefore, (27) becomes

$$\begin{aligned}
\mathcal{D}(\mathcal{N}_2 || \mathcal{N}_1) &= -\frac{1}{2} \ln \frac{|\mathbf{P}_2|}{|\mathbf{P}_1|} + \frac{1}{2} \int f(\mathbf{s}) \text{tr} \left[ \mathbf{P}_1^{-1} (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T \right] d\mathbf{s} \\
&\quad - \frac{1}{2} \int f(\mathbf{s}) \text{tr} \left[ \mathbf{P}_2^{-1} \mathbf{s} \mathbf{s}^T \right] d\mathbf{s}. \quad (32)
\end{aligned}$$

Since  $f(\mathbf{s})$  is scalar, it can be taken inside of the trace. Also, since the integral is a linear operator and the trace is essentially a summation, the order of the two can be interchanged.

Thus,

$$\begin{aligned}
\mathcal{D}(\mathcal{N}_2 || \mathcal{N}_1) &= -\frac{1}{2} \ln \frac{|\mathbf{P}_2|}{|\mathbf{P}_1|} + \frac{1}{2} \text{tr} \left[ \mathbf{P}_1^{-1} \int f(\mathbf{s}) (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1) (\mathbf{s} + \mathbf{x}_2 - \mathbf{x}_1)^T d\mathbf{s} \right] \\
&\quad - \frac{1}{2} \text{tr} \left[ \mathbf{P}_2^{-1} \int f(\mathbf{s}) \mathbf{s} \mathbf{s}^T d\mathbf{s} \right]. \quad (33)
\end{aligned}$$

After multiplying it out, the integral in the second term becomes

$$\begin{aligned}
& \int f(s)(s + \mathbf{x}_2 - \mathbf{x}_1)(s + \mathbf{x}_2 - \mathbf{x}_1)^T ds = \\
& \int f(s)ss^T ds + \int f(s)s(\mathbf{x}_2 - \mathbf{x}_1)^T ds + \int f(s)(\mathbf{x}_2 - \mathbf{x}_1)s^T ds + \int f(s)(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T ds \\
& = \int f(s)ss^T ds + \left( \int f(s)s ds \right)(\mathbf{x}_2 - \mathbf{x}_1)^T + (\mathbf{x}_2 - \mathbf{x}_1) \int f(s)s^T ds + (\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T \int f(s) ds. \quad (34)
\end{aligned}$$

Since  $f$  is the pdf of the random vector  $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_2)$ , it is even in  $s$ . Furthermore, both  $s$  and  $s^T$  are clearly odd in  $s$ , making both  $f(s)s$  and  $f(s)s^T$  odd functions in  $s$ . Therefore, since the limits of integration are over the whole space, we have

$$\int f(s)s ds = \int f(s)s^T ds = 0. \quad (35)$$

Noting that

$$\int f(s) ds = 1 \quad (36)$$

and that

$$\int f(s)ss^T ds = E_S \{ss^T\} = \mathbf{P}_2, \quad (37)$$

(34) simplifies to

$$\int f(s)(s + \mathbf{x}_2 - \mathbf{x}_1)(s + \mathbf{x}_2 - \mathbf{x}_1)^T ds = \mathbf{P}_2 + (\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T \quad (38)$$

and (33) simplifies to

$$\begin{aligned}
\mathcal{D}(\mathcal{N}_2 || \mathcal{N}_1) &= -\frac{1}{2} \ln \frac{|\mathbf{P}_2|}{|\mathbf{P}_1|} + \frac{1}{2} \text{tr} [\mathbf{P}_1^{-1} \mathbf{P}_2 + \mathbf{P}_1^{-1} (\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T] - \frac{1}{2} \text{tr} [\mathbf{P}_2^{-1} \mathbf{P}_2] \\
&= \frac{1}{2} [-\ln |\mathbf{P}_2 \mathbf{P}_1^{-1}| + \text{tr} [\mathbf{P}_1^{-1} \mathbf{P}_2] + \text{tr} [\mathbf{P}_1^{-1} (\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T] - \text{tr} [I]] \\
&= \frac{1}{2} [-\ln |\mathbf{P}_2 \mathbf{P}_1^{-1}| + \text{tr} [\mathbf{P}_1^{-1} \mathbf{P}_2] + \text{tr} [(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{x}_2 - \mathbf{x}_1)] - N],
\end{aligned}$$

where we have again used the cyclic property of the trace. Finally, observing that the quadratic form  $(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{x}_2 - \mathbf{x}_1)$  is a scalar,

$$\begin{aligned}
\mathcal{D}(\mathcal{N}_2 || \mathcal{N}_1) &= \frac{1}{2} [-\ln |\mathbf{P}_2 \mathbf{P}_1^{-1}| + \text{tr} [\mathbf{P}_1^{-1} \mathbf{P}_2] + (\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{P}_1^{-1} (\mathbf{x}_2 - \mathbf{x}_1) - N] \\
&= \frac{1}{2} [(\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{P}_1^{-1} (\mathbf{x}_1 - \mathbf{x}_2)] + \frac{1}{2} [\text{tr} [\mathbf{P}_1^{-1} \mathbf{P}_2] - \ln |\mathbf{P}_2 \mathbf{P}_1^{-1}| - N].
\end{aligned}$$

## REFERENCES

- [1] Y. Bar-Shalom, "Tracking methods in a multitarget environment," *IEEE Trans. on Automatic Control*, vol. AC-23, pp. 618–626, Aug. 1978.
- [2] P. Smith and G. Buechler, "A branching algorithm for discriminating and tracking multiple objects," *IEEE Trans. on Automatic Control*, vol. AC-20, pp. 101–104, Feb. 1975.
- [3] O. E. Drummond, "Best hypothesis target tracking and sensor fusion," in *Signal and Data Processing of Small Targets* (O. E. Drummond, ed.), vol. SPIE Proc. 3809, (Denver, CO), pp. 586–600, July 1999.
- [4] G. W. Pulford, "Taxonomy of multiple target tracking methods," *IEE Proc. F: Radar, Sonar, and Navigation*, vol. 152, pp. 291–304, Oct. 2005.
- [5] R. A. Singer, R. G. Sea, and K. B. Housewright, "Derivation and evaluation of improved tracking filters for use in dense multitarget environments," *IEEE Trans. on Information Theory*, vol. IT-20, pp. 423–432, July 1974.
- [6] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, 1995.
- [7] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. on Automatic Control*, vol. AC-24, pp. 843–854, Dec. 1979.
- [8] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [9] I. J. Cox and S. L. Hingorani, "An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 138–150, Feb. 1996.
- [10] K. G. Murty, "An algorithm for ranking all the assignments in order of increasing cost," *Operations Research*, vol. 16, pp. 682–687, 1968.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [12] T. Kirubarajan and Y. Bar-Shalom, "Kalman filter versus imm estimator: When do we need the latter?," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, pp. 1452–1457, Oct. 2003.
- [13] E. Gul, "On the track similarity test in track splitting algorithm," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 30, pp. 604–606, Apr. 1994.

- [14] S. Kullback, *Information Theory and Statistics*. John Wiley & Sons, Inc., 1959.
- [15] M. Nakao, M. Sawada, Y. Matsushita, N. Katayama, M. Yamamoto, and M. Munakata, “Characterization of multivariate cardiovascular dynamics during head-up tilting based on kullback-leibler divergence,” *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2180–2183, 1997.